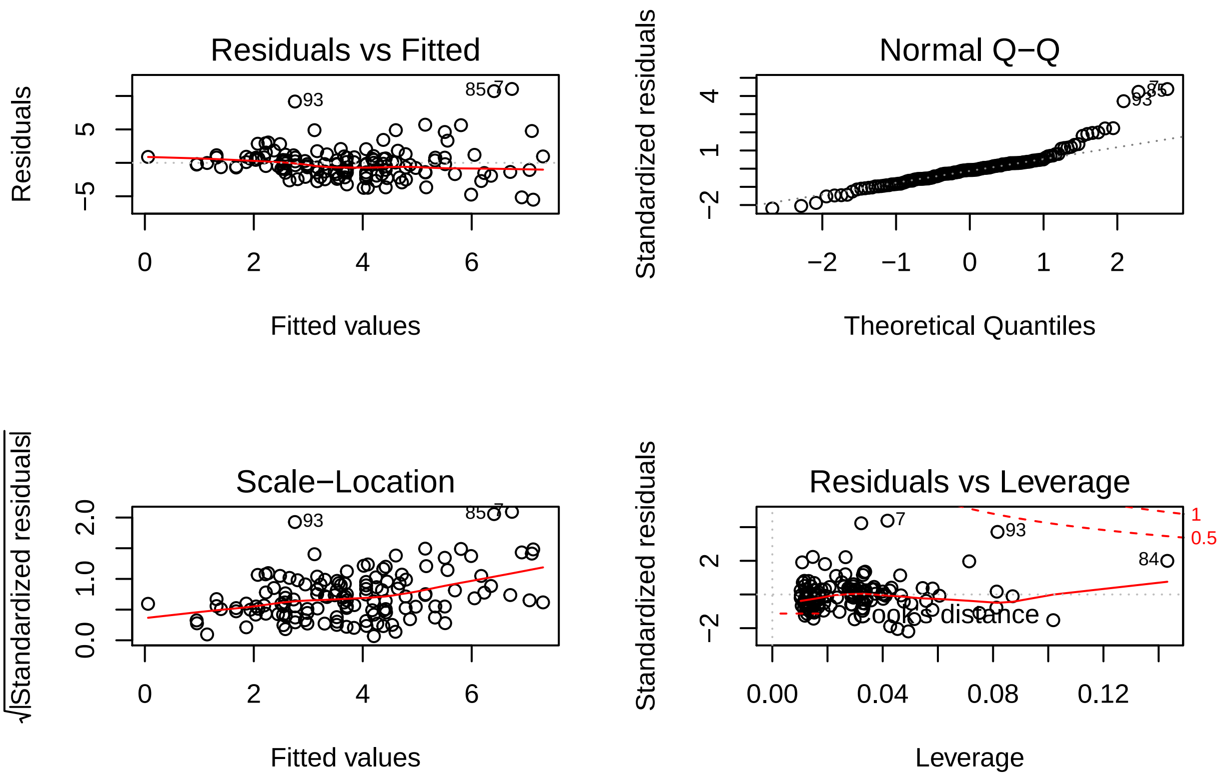
yang_seonhyeHW18

```
library(data.table)
library(MASS)
library(readr)
```

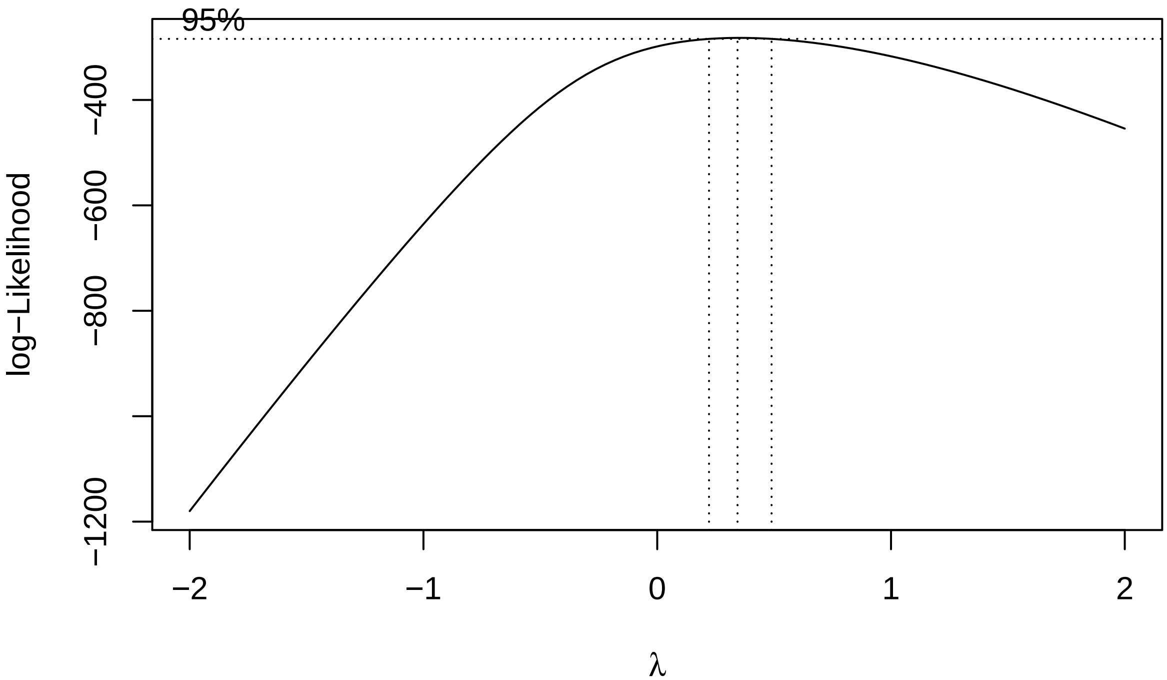
```
data<- fread("http://users.stat.ufl.edu/~winner/data/fishermen_mercury.csv")
fishermen_mercury<- data.table("fishermen_mercury.csv", header = T)
```

Question 1

```
attach(data, warn.conflicts = F)
fit<- lm(TotHg~fisherman+height+weight)
par(mfrow=c(2, 2))
plot(fit)
```



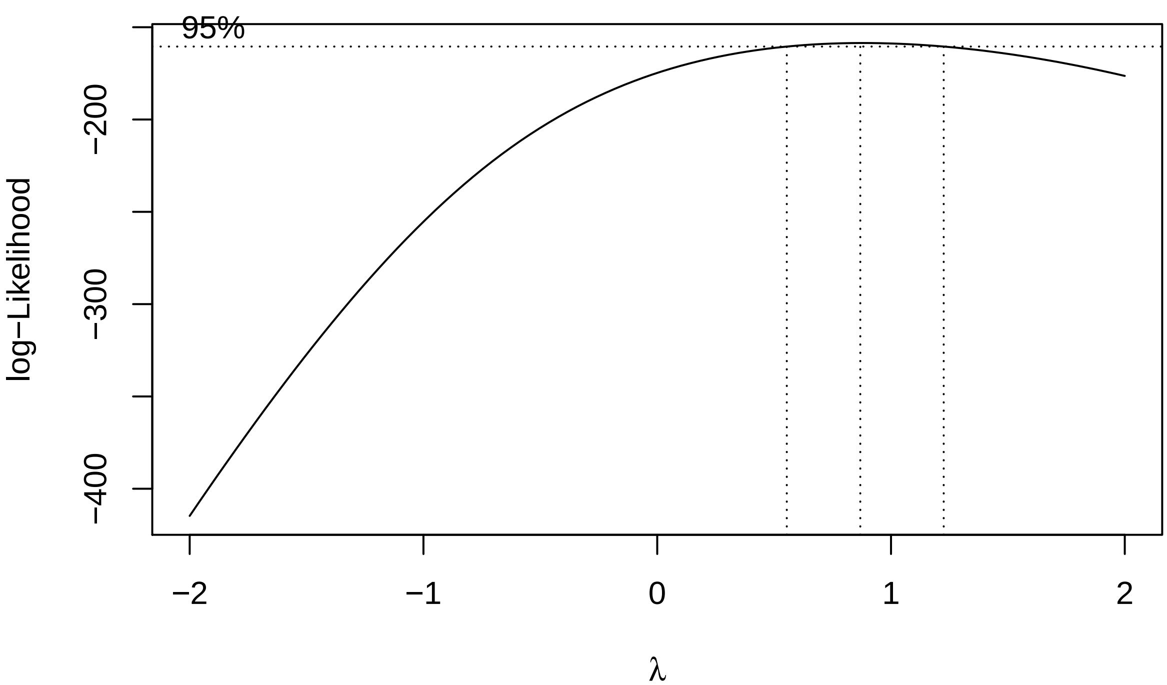

```
boxcox(fit)
```



Looking at the Residuals VS Fitted graph, there is a nearly straight trendline and also we can see that the residuals are mostly grouped together, indicating that this is a good model. The Normal Q-Q plot is also linear, but the ending points are deviating from this, indicating problems with the model. Finally, the Scale-Location plot has a linear trendline, but not a flat one, and the Residuals VS Leverage has some far outling points. Combining these metrics we can say our model is satisfactory, but can be improved.

Question 2

```
fit2<- lm(TotHg^(2/5)~fisherman+height+weight)
boxcox(fit2)
```

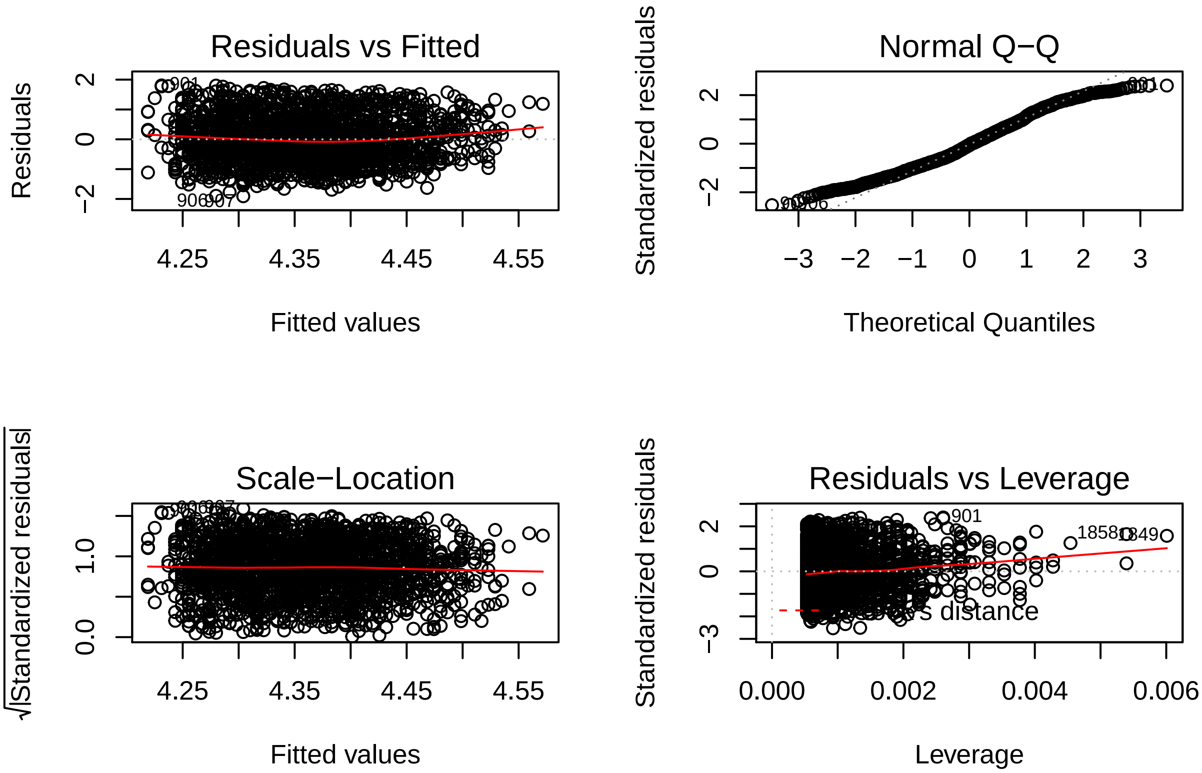


$y^{2/5}$ is appropriate because we can see that lambda is much closer to 1. Where $\lambda = 2/5$ because we know that $\lambda = 1$ will just give us the the unfitted model.

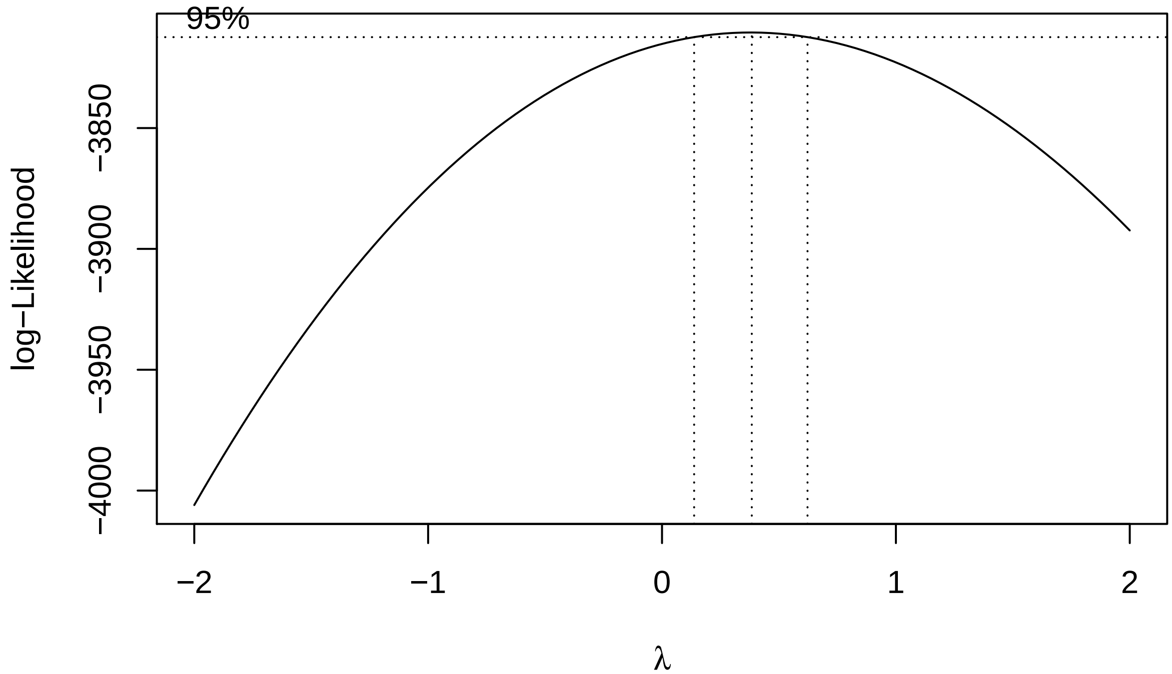
```
data1<- fread("http://users.stat.ufl.edu/~winner/data/napa_marathon_fm2015.csv")
napa_marathon_fm2015<- data.table("napa_marathon_fm2015", header = T)
```

Question 3

```
attach(data1, warn.conflicts = F)
fit3<- lm(Hours~Age)
par(mfrow=c(2, 2))
plot(fit3)
```



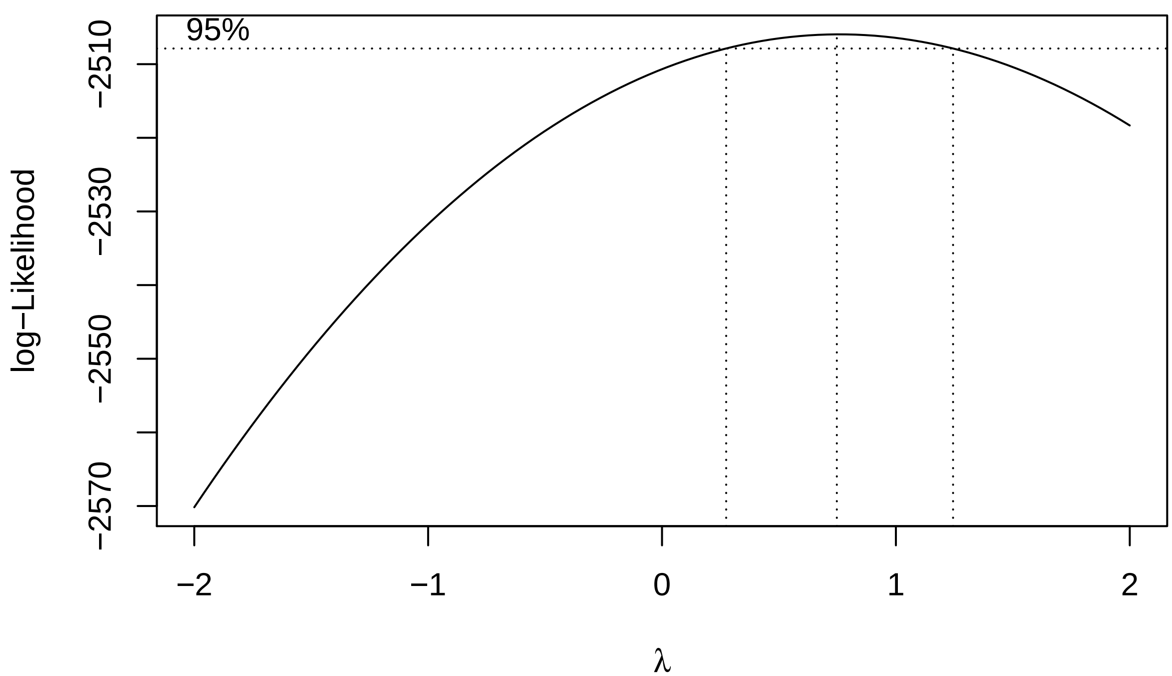

```
boxcox(fit3)
```



Looking at the Residuals VS Fitted plot, we have a flat trendline and evenly distributed points, indicating that our model had decent fit. Out Normal Q-Q plot is very linear with few points being far out of place. Finally, with our Scale-Location and Residual Vs Leverage Plots, we have linear and flat trendlines, and our points are properly spaced, indicating our model has a good fit.

Question 4

```
attach(data1, warn.conflicts = F)
fit4<- lm(Hours^(1/2)~Age)
boxcox(fit4)
```



Looking at the adjusted, there is a difference than before. We can see that lambda is closer to 1 than the previous model.

Question 5

L'hopital's rule is defined as

$$\lim(xß0) = \frac{f'(x)}{g'(x)}$$

This is Box-cox transformation:

$$B(x, \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0, \\ \ln(x), & \text{if } \lambda = 0. \end{cases} \tag{1}$$

And lastly, this is what we want to find:

$$\lim_{\lambda \to 0} \frac{y^\lambda - 1}{\lambda} = \ln(y)$$

First, we mulitply the numerator by e:

$$y^\lambda = e^{y^\lambda} = e^{\lambda \ln(y)}$$

and we can ignore the constant because we just need to dominant terms. Now applying l'hopital rule:

$$\lim(xß0) = \frac{f'(x)}{g'(x)}$$

$$\lim(\lambda ß0) = \frac{f'(\lambda)}{g'(\lambda)}$$

$$\lim(\lambda ß0) \frac{\frac{d(e^{\lambda \ln(y)})}{d\lambda}}{\frac{d(\lambda)}{\lambda}} = \frac{\ln(y) e^{\lambda \ln(y)}}{1}$$

As

$$\lambda ß0 = \ln(y)$$